A Toolkit for Short and Noisy Time Series

The SSA-MTM Toolkit

Michael Ghil* and Pascal Yiou**

*ENS, Paris, and UCLA

**LSCE, Gif-sur-Yvette

A Free Toolkit for Spectral Analysis

- The SSA-MTM Toolkit:
- Developed under the supervision of M. Ghil (UCLA) since 1994.
- GUI-based, adapted to linux, unix and MacOSX platforms.
- Latest developments by D. Kondrashov (UCLA).
- Available at: www.atmos.ucla.edu/tcd/ssa

Features

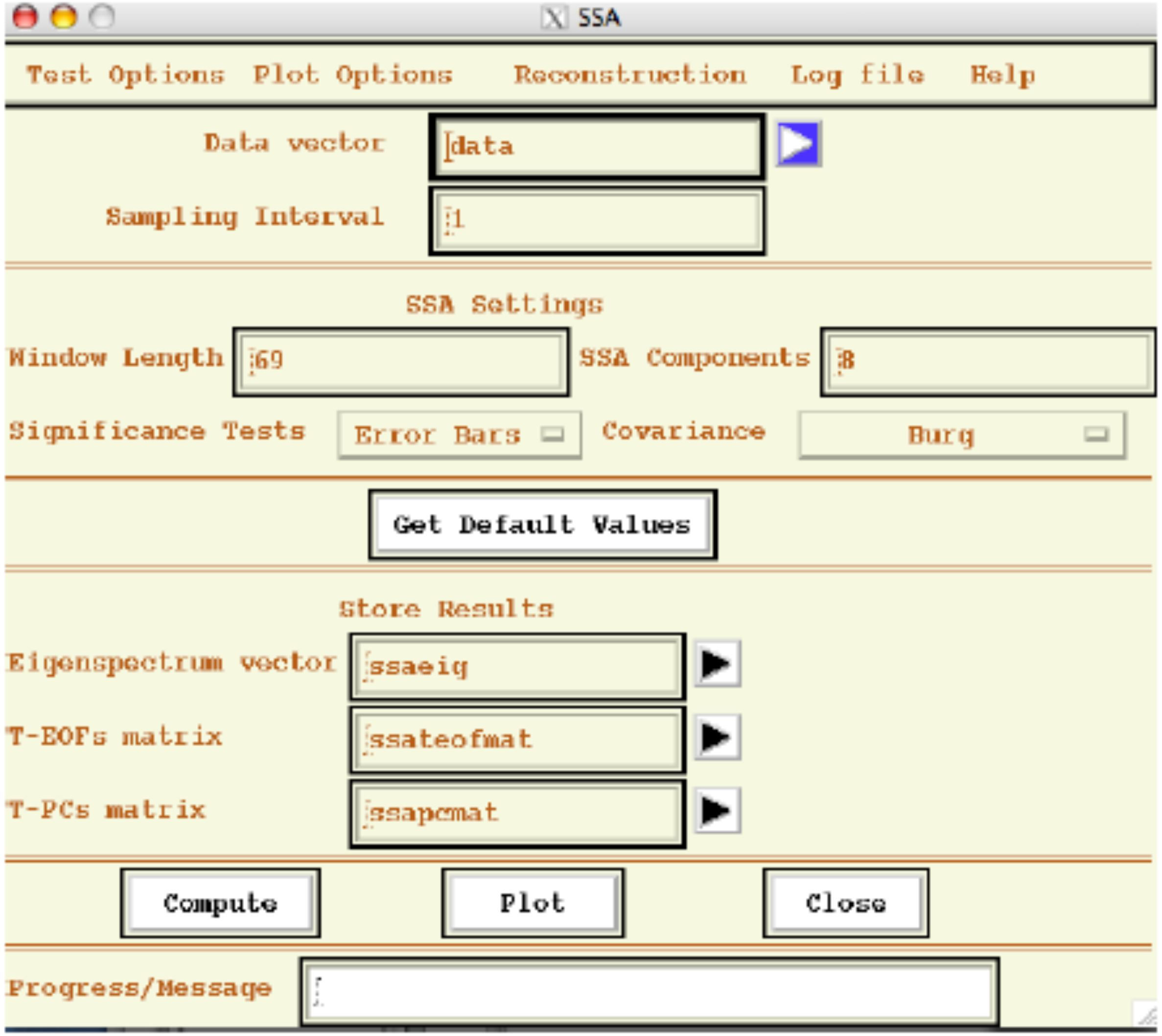
- Graphics interface with xmgrace & idl.
- Online documentation
- Contains:
  - Blackman-Tukey (BT) method
  - Maximum entropy method (MEM)
  - Multi-Taper Method (MTM)
  - Singular Spectrum Analysis (SSA)

General Goals

- Reduce the variance of the spectral estimate of a time series based on the periodogram (MTM), correlogram (BT) or other (SSA).
- Estimate peak frequencies fingerprinting of limit cycles of the underlying dynamical system.
- Provide confidence intervals when such behavior is blurred by noise.

Targeted audience

- Non-specialists in time series analysis
  - Reasonable default options
  - Reads ASCII files
- Non-specialists in computer management
  - Precompiled binaries
  - User-friendly interface



Type of noise used in the toolkit

- Red noise:
  - AR(1) random process: $X(t+1) = aX(t) + b(t), 0 < a < 1$
  - Continuous spectrum with negative slope (due to thermal or mechanical inertia)

$$C_X(\tau) = \frac{\sigma^2 a^{|\tau|}}{1 - a^2}$$

$$P_X(f) = C_X(0) \frac{1 - a^2}{1 - 2a\cos 2\pi f + a^2}$$

Blackman-Tukey method (BT: 1958)

The variance of the periodogram estimate of the spectrum is proportional to N, the number of points in a time series.

BT: use the correlogram to weigh the high-rank auto-correlations (i.e., high frequencies) by weights $w_m(t)$ (or *tapers*):

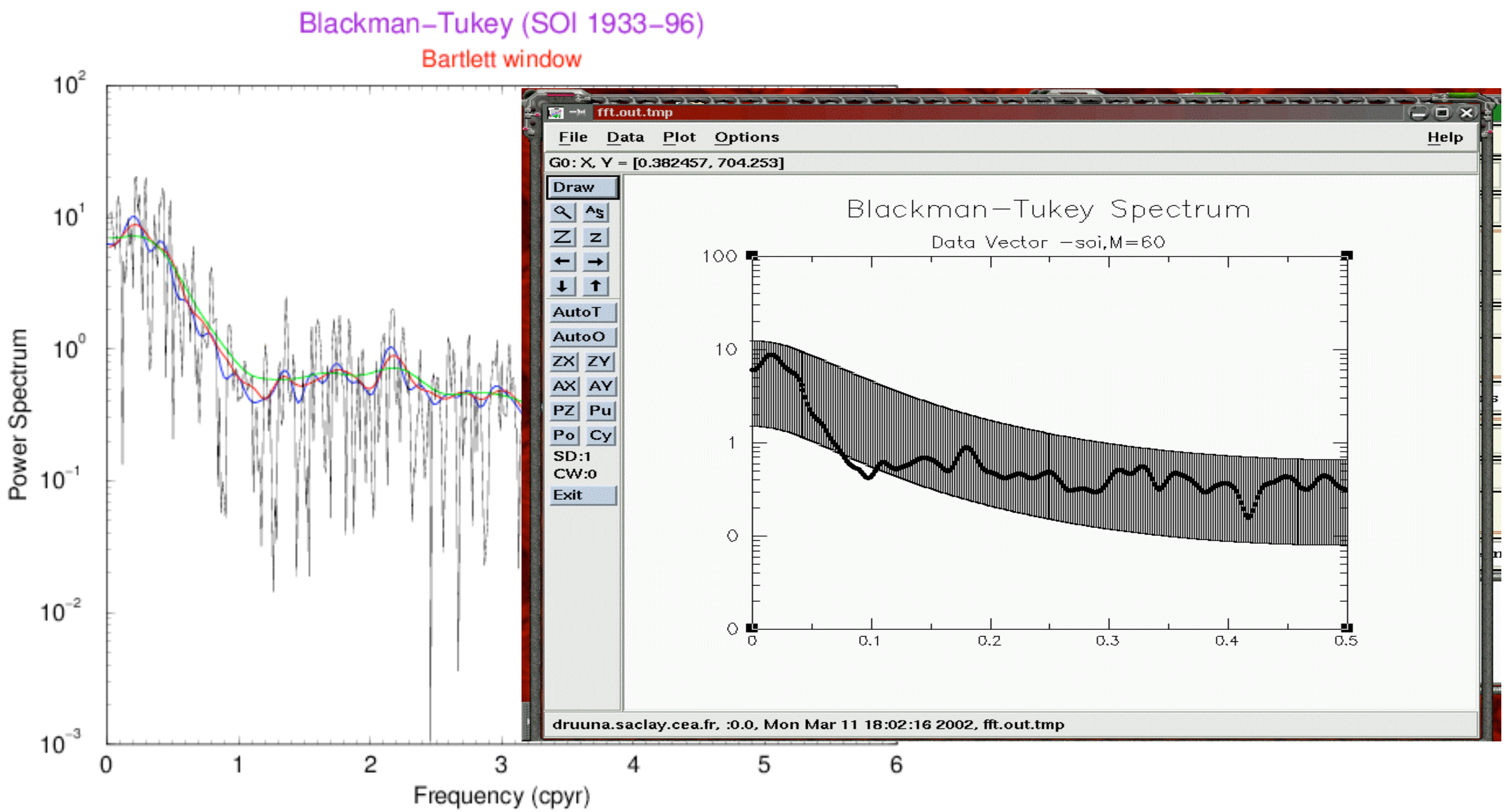
$$\tilde{P}_X^{(2)}(f) = \sum_{t=-(M-1)}^{M-1} w_m(t) C_X(t) e^{-i2\pi f t}$$

BT method consists in choosing $w_m(t)$ and the maximum $M < N$ on which the taper acts (“*window carpentry*” + “*window opening and closing*”)

Properties of BT

- Efficient estimate of continuous parts of the spectrum because of the smoothing.
- Fast computation.
- Not very useful for discrete parts of the spectrum (poor frequency resolution).
- Heuristic choice of order of auto-correlation for the correlogram.
- Heuristic carpentry of tapers, typically “triangles” or “cosinuses”.

An example (SOI time series)



Estimated variance is reduced, but huge loss of frequency resolution.

Multi-Taper Method (MTM: Thomson, 1982)

Determine K uncorrelated estimates of the periodogram from well chosen tapers $w_k(t)$, that minimize spectral leaks (Slepian, 1978).

$$P_X^{(k)}(f) = \left| \sum_{t=1}^{N} w_k(t) X(t) e^{-i2\pi f t} \right|^2, k = 1 \ldots K$$

The multitaper estimate is a weighted average of the K spectra:

$$\tilde{P}_X(f) = \frac{\sum_{k=1}^{K} \lambda_k P_X^{(k)}(f)}{\sum_{k=1}^{K} \lambda_k}$$

This estimate achieves a compromise between high frequency resolution for periodic signals, and low variance.

MTM tests

Harmonic analysis: estimate periodic components and their amplitude

$$X(t) = \mu e^{i2\pi f_0 t} + b(t)$$

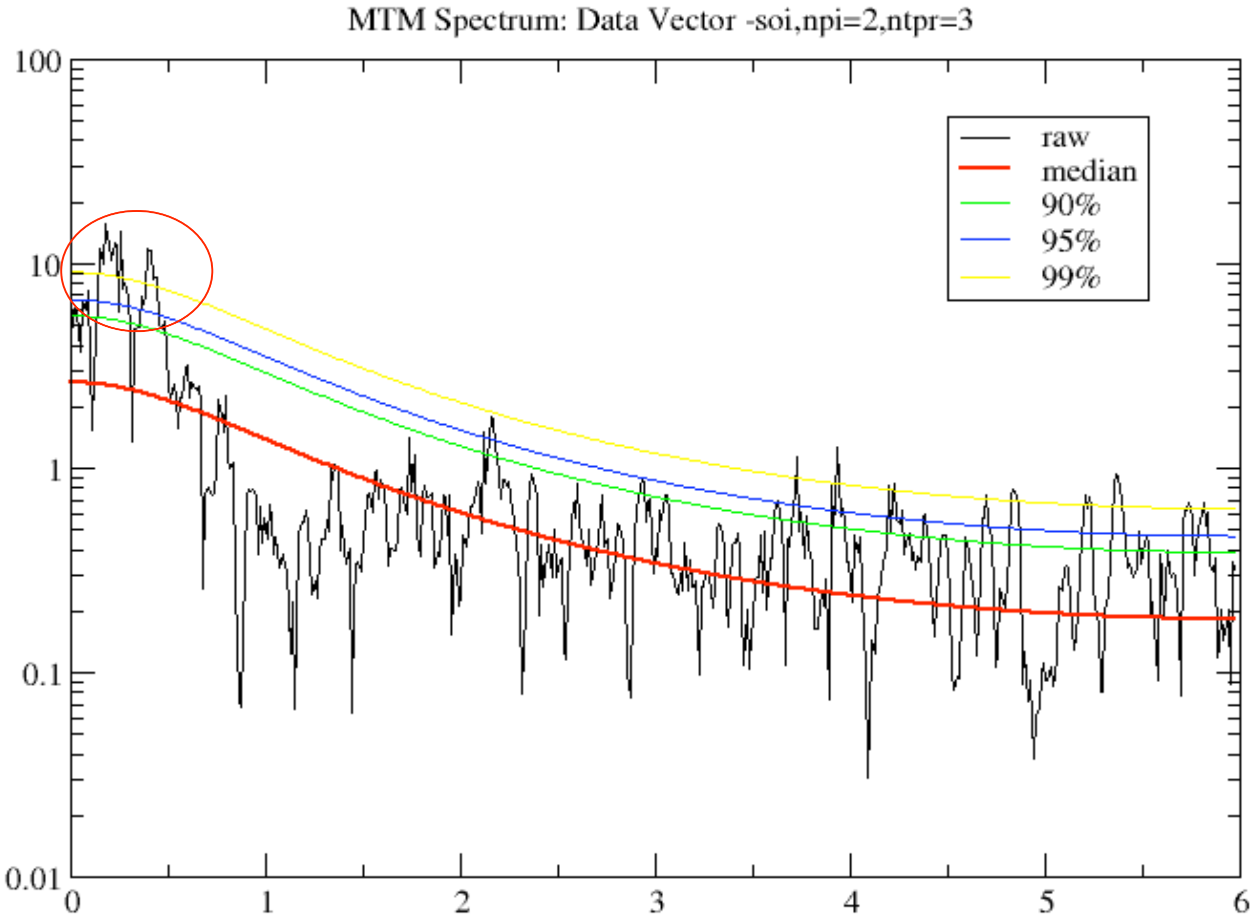
The parameters μ and f_0 are estimated by a least-square fit, and this model is tested with a Fisher-Snedecor test against white noise.

Red-noise test (Mann & Lees, 1996): median smoothing of the spectrum to obtain the “equivalent” red noise of the data, and the distribution of its spectrum. Useful to detect non-harmonic outliers.

Features of MTM

- Efficient in detecting periodic (“line”) components when they do exist.
- A random signal can generate many (falsely) “significant” peaks.
- Two ways of testing the spectrum (harmonic and red-noise tests)

Application to SOI time series

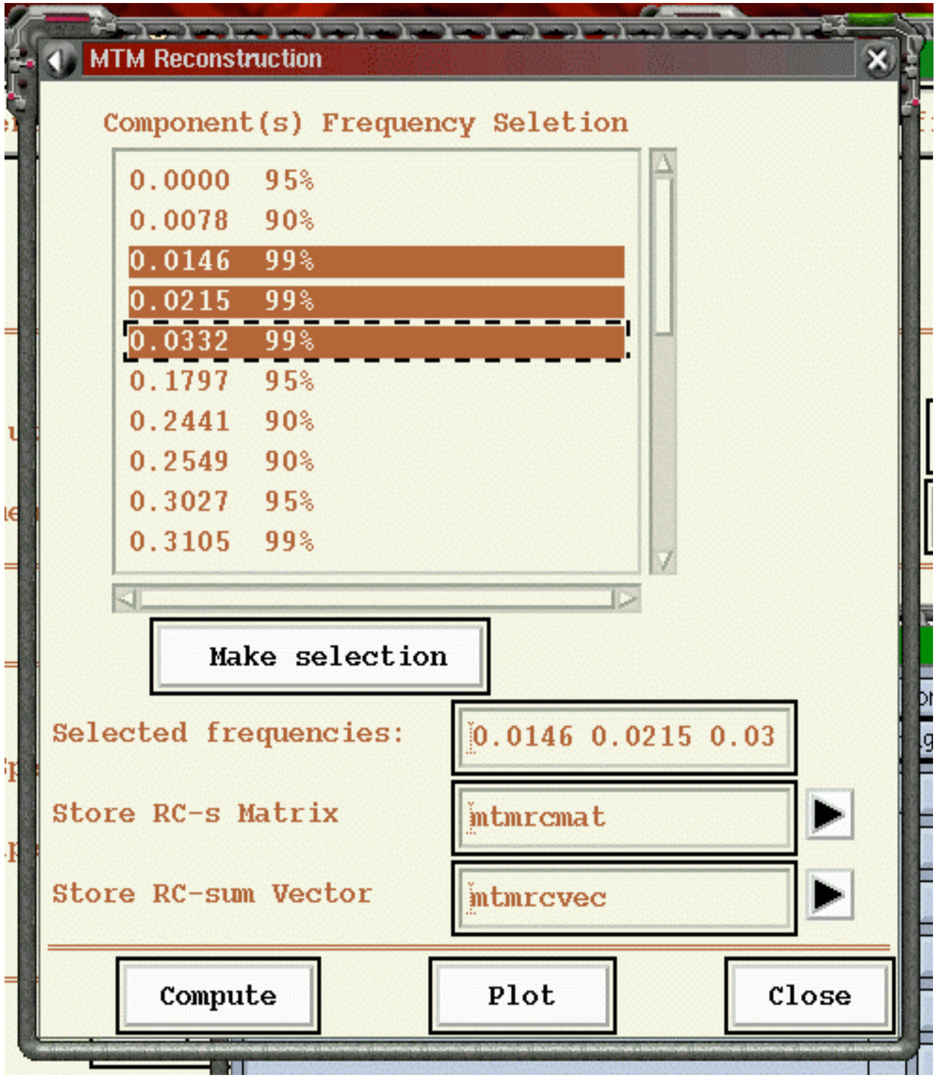


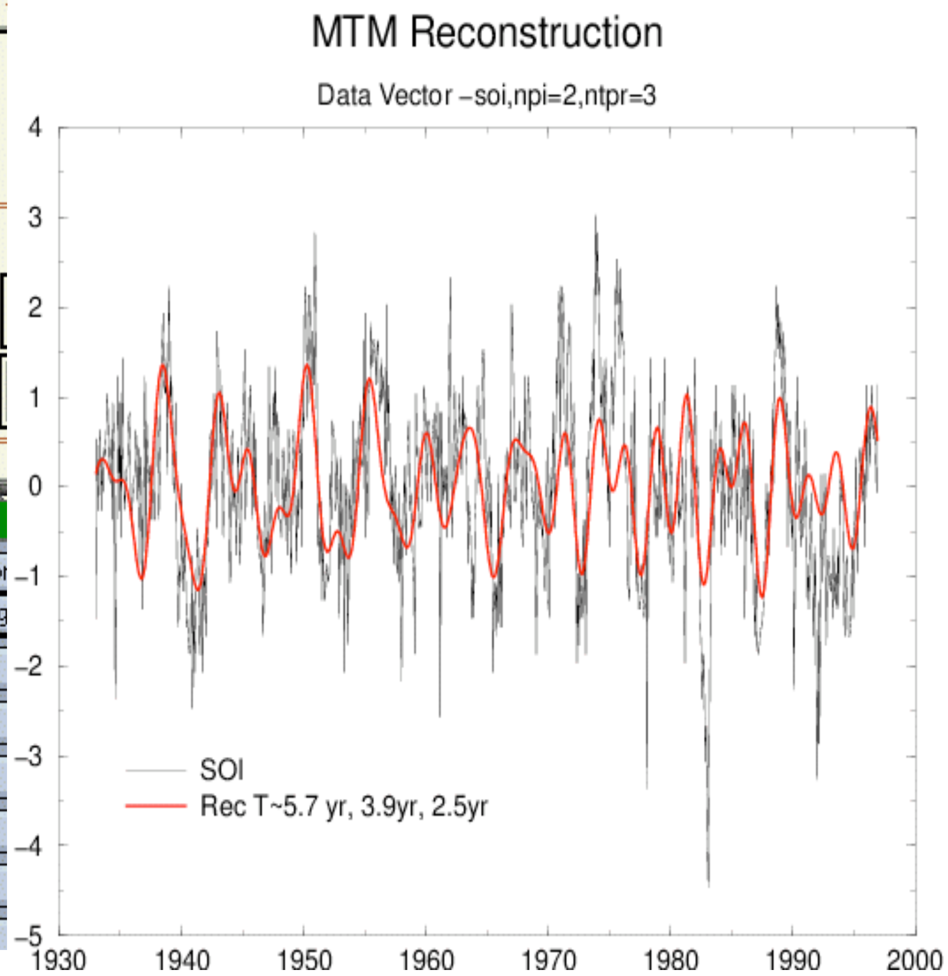
MTM filtering and reconstruction

- When "peaks" are detected, it is possible to reconstruct the corresponding time series.
- This can be done by taking into account the width of the spectral peak.

$$\tilde{X}(n\Delta t) = \Re\left\{ A_n e^{-2\pi i f_0 n\Delta t} \right\}.$$

MTM Reconstruction





Singular Spectrum Analysis (SSA)

$X(t)$, t=1…, N: Time series of an observable of a dynamical system

M: embedding dimension

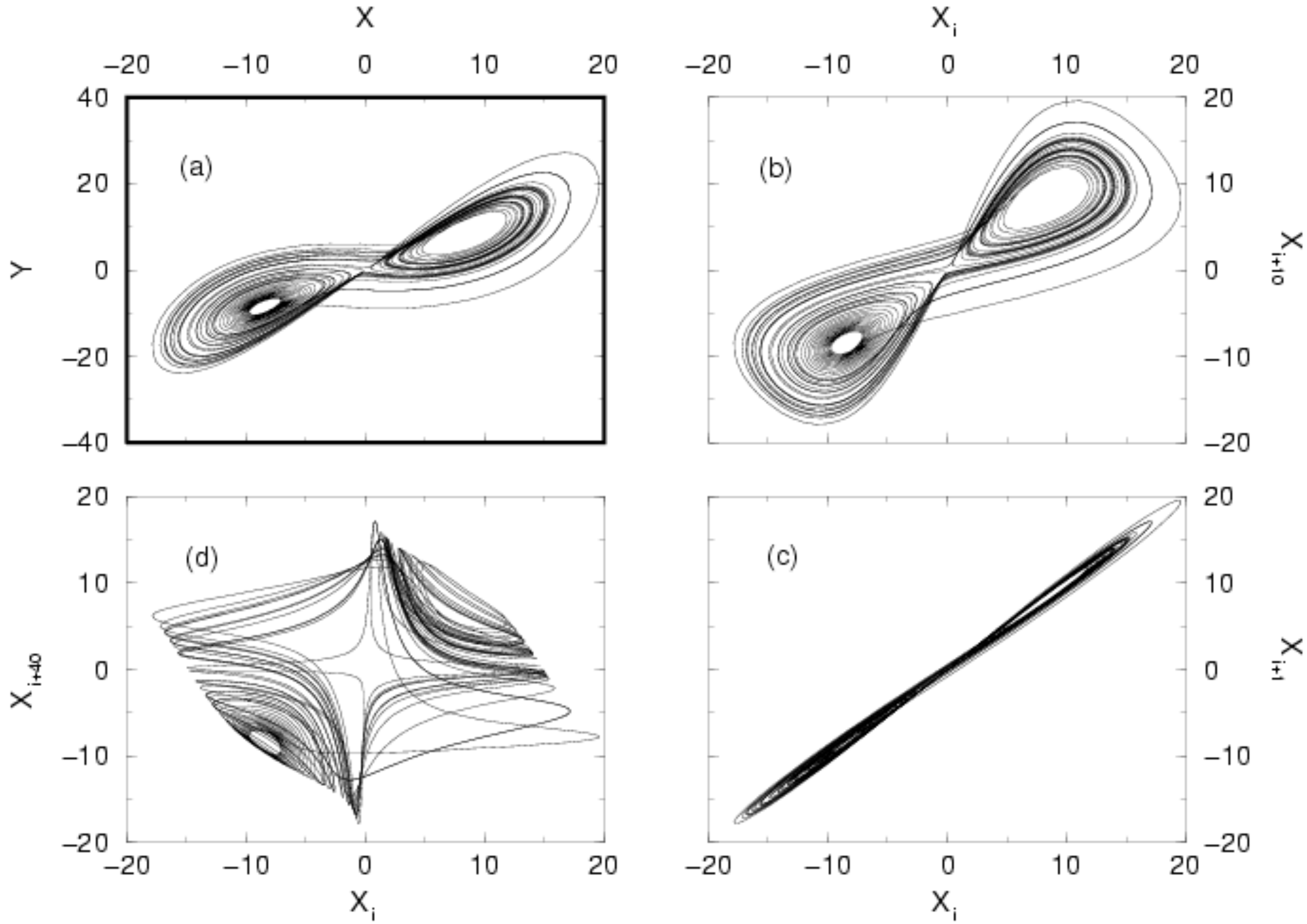
Trajectory of X in dimension M:

$$\Xi = \begin{pmatrix} X(1) & \cdots & X(N-M+1) \\ \vdots & & \vdots \\ X(M) & \cdots & X(N) \end{pmatrix}$$

Trajectory is visible in dimension $M \leq 3$, but looks like a (multi-dimensional) spaghetti dish otherwise!

SSA Motivation

Mañe-Takens theorem (1981) on attractor reconstruction, based on the "method of delays."



Covariance matrix C_Ξ of Ξ (*moments of inertia*) and its eigenelements

Principal directions of Ξ:

$$C_\Xi \, \rho_k = \lambda_k \rho_k$$

Eigenvectors (*empirical orthogonal functions*, EOFs)

Variance of mode *k*

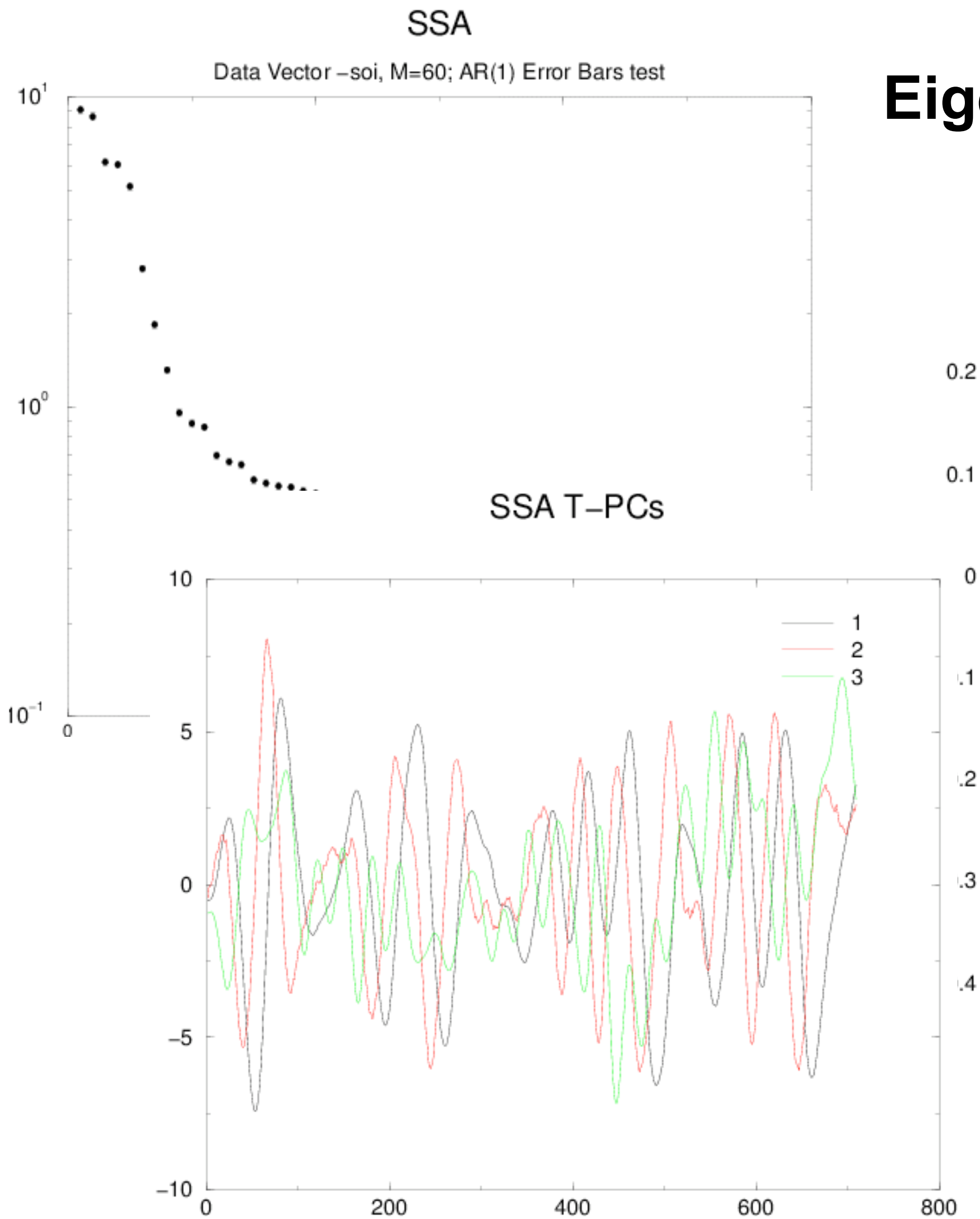
Principal components (PCs): Projections of X onto EOFs

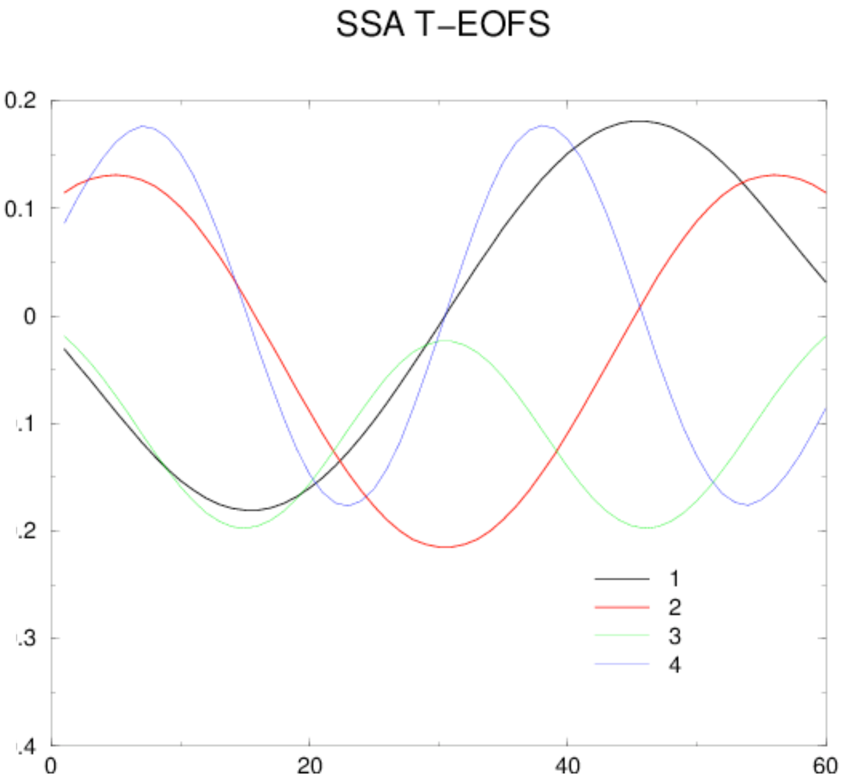
$$X_k(t) = \sum_{j=1}^{M} X(t+j)\rho_k(j)$$

Reconstruction from a subset K of components (filtering):

$$\widehat{X}_K(t) = \frac{1}{A} \sum_{k \in K} \sum_{j=1}^{M} \rho_k(j)\, X_k(t-j), \quad M \le t \le N - M + 1$$

Eigenvalues and eigenvectors:





Monte Carlo SSA

(Allen & Smith, *J. Clim.*, 1995)

Goal: Assess whether the SSA spectral estimation can reject the null hypothesis that the time series is red noise.

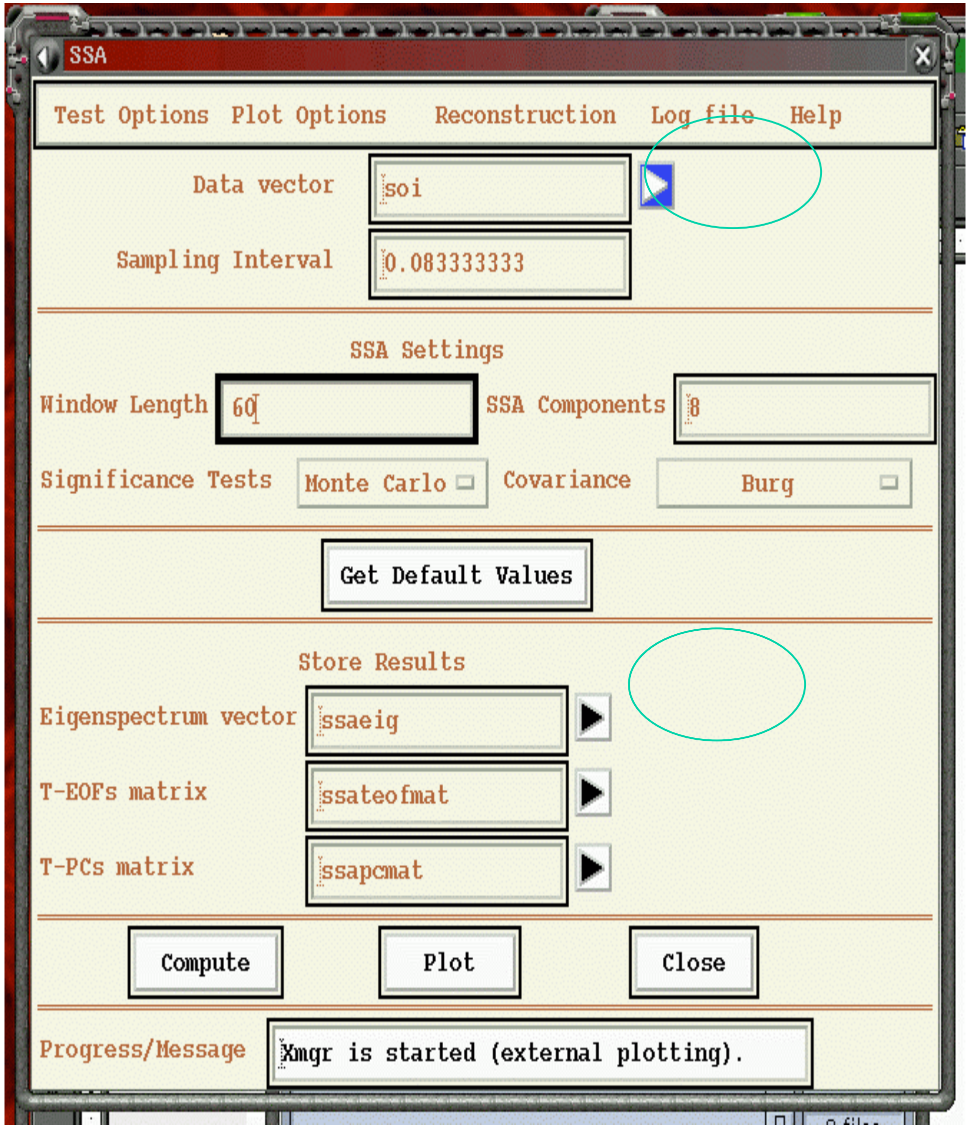
Procedure:

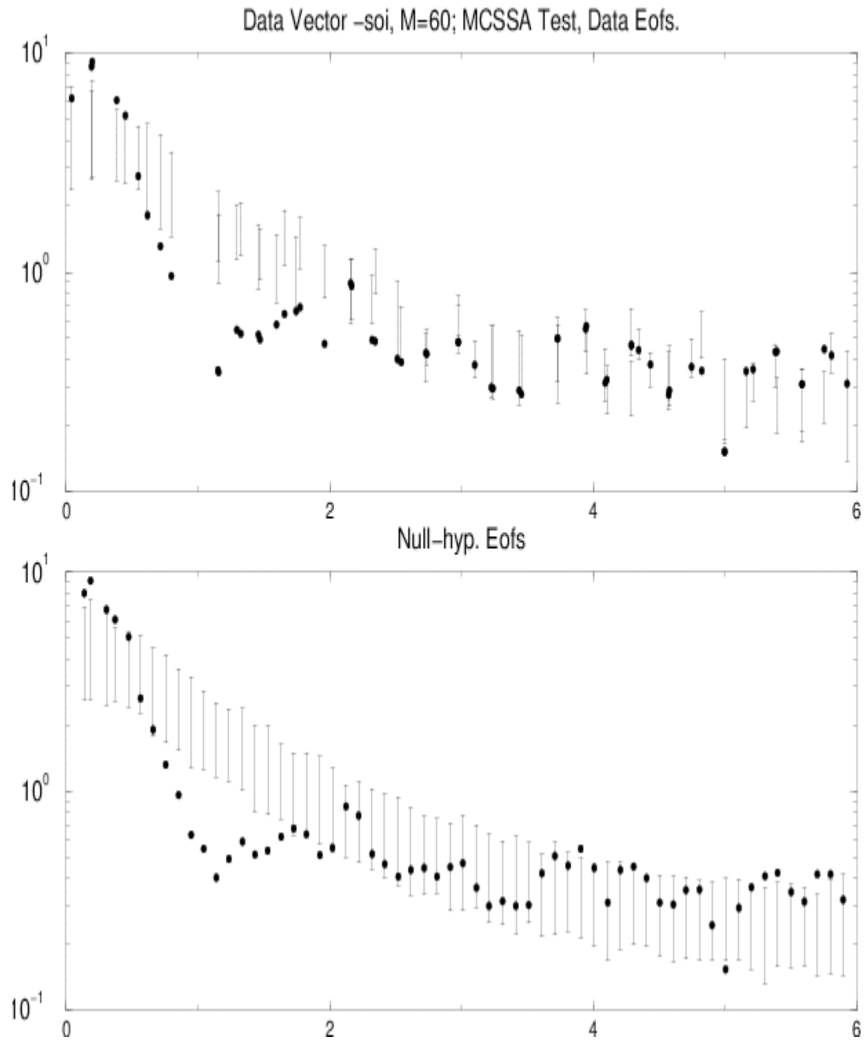
- Estimate the parameters of a red-noise process with the same variance and lag-covariance as the observed time series $X(t)$.
- Compare the pdf of the projection of the noise covariance onto the data EOFs:

$$\Lambda_B = R_X^t \bullet \underbrace{C_R}_{\text{Covar. red noise}} \bullet \overbrace{R_X}^{\text{EOFs data}}$$

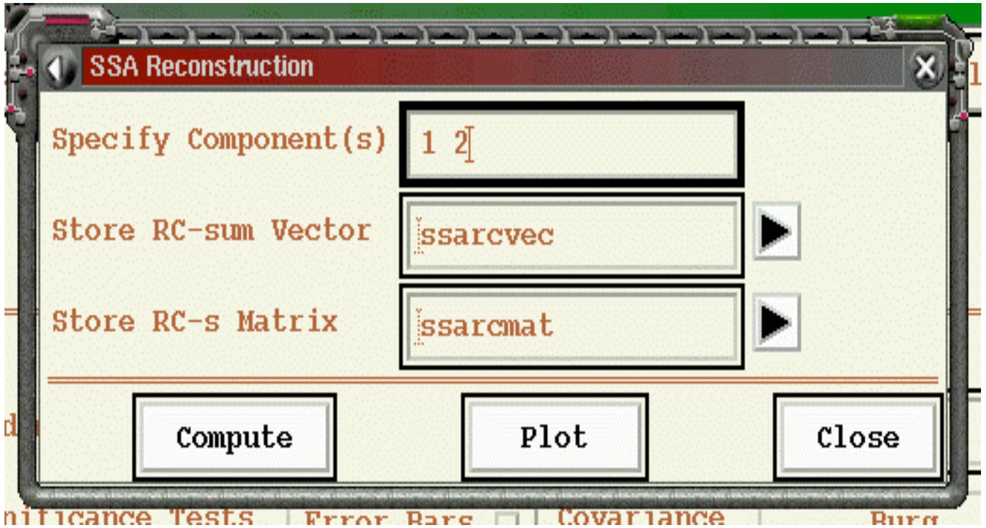
The null hypothesis is rejected using the pdf of Λ_B.

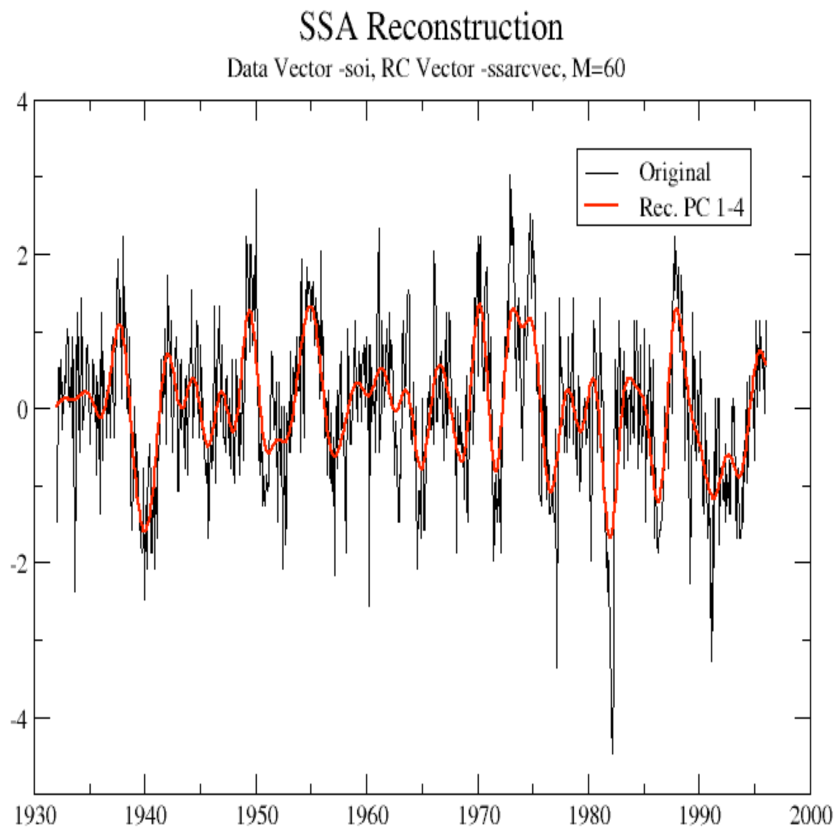
Monte Carlo SSA: red-noise test





Component reconstruction





Conclusions

- The toolkit presents an array of techniques to estimate the spectrum of a time series.
- These techniques make different underlying assumptions about the process generating the time series (stationarity, normality, etc.).
- Using several methods allows one to check for the robustness of the results (e.g. peak frequency estimates).
- The tests largely use an AR(1) null hypothesis.

Future work

- General need for preprocessing data (with R, matlab, excel…)
- Orientation towards “beginners” (click-O-drome interface)
  - Need for automated or batch use for “pros”
- R version of the toolkit with LRP features and a universal (incl. Windows) interface?
  - Use of already existing code in C and Fortran

References

- C. Chatfield, *The Analysis of Time Series: An Introduction*, Chapman and Hall, New York, 1984
- M. Ghil *et al.*, Advanced spectral methods for climatic time series, *Rev. Geophys.*, 40, doi:10.1029/2001RG000092, 2002
- D. B. Percival et A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, 1993
- http://www.ipsl.jussieu.fr/CLIMSTAT/
- http://www.atmos.ucla.edu/tcd/ssa
- http://www.r-project.org